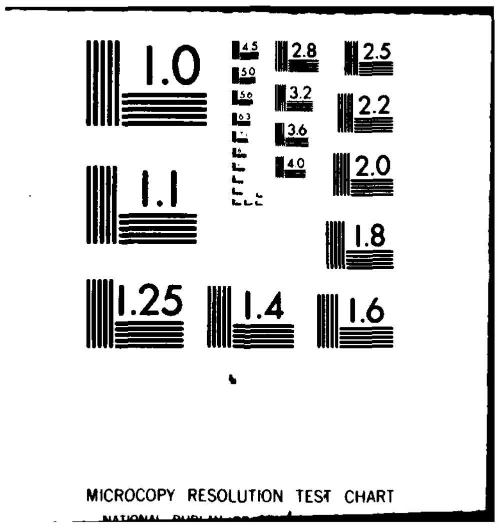
AD-A089 656 FLORIDA STATE UNIV TALLAHASSEE DEPT OF STATISTICS F/G 12/1
SMALL SAMPLE PROPERTIES OF TWO SURVIVAL FUNCTION ESTIMATORS BAS--ETC(U)
JUL 80 Y CHEN, M HOLLANDER, N A LANGBERG
UNCLASSIFIED FSU STATISTICS M550 AFOSR-TR-80-0931 NL

1 of 1

END
DATE FILMED
10-80
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
MICROCOPY RESOLUTION TEST CHART

(6) Small Sample Properties of Two Survival Function Estimators Based on Incomplete Data

by

(10) Yuan-Yan Chen[1], Myles Hollander[1] and Naftali A. Langberg[1]

(14) FSU Statistics [redacted] No. M550, TR-78-112
AFOSR Technical Report No. 78-112

(9)

(11) Jul[redacted]80 (12) 21

The Florida State University
Department of Statistics
Tallahassee, Florida 32306

(15) AFOSR-78-3678 (16) 23p

(17) A5

[1]Research supported by the Air Force Office of Scientific Research, AFSC, USAF, under Grant AFOSR-78-3678.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 80-0931	2. GOVT ACCESSION NO. AD-A089656	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SMALL SAMPLE PROPERTIES OF TWO SURVIVAL FUNCTION ESTIMATORS BASED ON INCOMPLETE DATA		5. TYPE OF REPORT & PERIOD COVERED Interim
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Yuan-Yan Chen, Myles Hollander, Naftali, A. Langberg		8. CONTRACT OR GRANT NUMBER(s) AFOSR 78-3678
9. PERFORMING ORGANIZATION NAME AND ADDRESS Florida State University, Department of Statistics, Tallahassee, FL 32306		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332		12. REPORT DATE July 1980
		13. NUMBER OF PAGES 18
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Product limit estimator, Piecewise exponential estimator, survival function, proportional hazards, right-censoring, small sample comparisions.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

For estimating an underlying survival distribution, we consider two estimators based on a randomly right-censored sample: The traditional Product Limit Estimator (PLE), introduced by Kaplan and Meier (1958), and a competitor, the Piecewise Exponential Estimator (PEXE), introduced by Kitchin, Langberg, and Proschan (1980a). Under a proportional hazards model we present formulas for the mean and variance of the PLE and find upper and lower bounds for mean and variance of the PEXE. These new expressions for finite sample sizes are compared with known asymptotic results.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Small Sample Properties of Two Survival Function Estimators Based on Incomplete Data

by

Yuan-Yan Chen, Myles Hollander and Naftali A. Langberg

Abstract.

For estimating an underlying survival distribution, we consider two estimaters based on a randomly right-censored sample: The traditional Product Limit Estimator (PLE), introduced by Kaplan and Meier (1958), and a competitor, the Piecewise Exponential Estimator (PEXE), introduced by Kitchin, Langberg and Proschan (1980a). Under a proportional hazards model we present formulas for the mean and variance of the PLE and find upper and lower bounds for mean and variance of the PEXE. These new expressions for finite sample sizes are compared with known asymptotic results.

Key words: Product limit estimator, Piecewise exponential estimator, survival function, proportional hazards, right-censoring, small sample comparisions.

1. INTRODUCTION

Let X_1, X_2, ... be independent identically distributed (i.i.d.) random variables (r.v.'s) denoting lifelengths with a common continuous distribution function (d.f.) F, and let Y_1, Y_2, ... be i.i.d. right-censoring r.v.'s with a common continuous d.f. H. We assume that (X_1, Y_1), (X_2, Y_2), ... is an i.i.d. sequence of random pairs with independent nonnegative components defined on a common probability space.

Let $I(A)$ denote the indicator function of the set A, let $Z_i = \min\{X_i, Y_i\}$ and $\delta_i = I(X_i \leq Y_i)$, $i = 1, \ldots, n$. We consider the problem of estimating $\bar{F} \equiv 1 - F$, the underlying survival function, from the sample $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$.

Kitchin, Langberg, and Proschan (KLP)(1980a) introduce a new estimator of $\bar{F}$; the <u>Piecewise Exponential Estimator</u> (PEXE), a competitor to the traditional <u>Product Limit Estimator</u> (PLE), introduced by Kaplan and Meier (1958). KLP show that the PEXE is a strongly consistent estimator of $\bar{F}$ and that the standarized PEXE process converges weakly to a Gaussian process. This Gaussian process has the same covariance structure as that of the standarized PLE process. [For details see KLP (1980a) and (1980b).] Thus the PEXE has the same asymptotic behavior as the PLE. [For the strong consistency of the PLE see Peterson (1977), Langberg, Proschan, and Quinzi (1980) or KLP (1980a). For the weak convergence of the standarized PLE see Breslow and Crowley (1974) or KLP (1980b).] In contrast to the PLE, which is a step function with jumps at the observed failures, the PEXE is a continuous and strictly decreasing function up to the last failure. In many life testing situations the survival function is anticipated to be strictly decreasing smoothly over time. Thus, in many situations the PEXE will be the more appropriate than the PLE. Further, the value of the PEXE in an interval

between any two successive failures depends on the actual withdrawal times up to the failure time that determines the right-hand side of the interval, and on that failure time. In contrast the value of the PLE between two successive failures depends on the number of withdrawals up to the failure that determines the left-hand side of the interval, and does not depend on the right-hand side failure. [For further details see KLP (1980a).]

In this paper we consider the small sample properties of the two estimators. Our calculations are performed under the assumption that the life-length X and the time to censorship Y have proportional hazards. (See Definition 2.3.)

In Section 2 we define the PEXE, the PLE and the proportional hazards models. Then we present a useful characterization of these models that is used in the last two sections.

In Section 3 we present formulas for the mean and variance of the PLE, and upper and lower bounds for the mean and variance of the PEXE. All computations in Section 3 are conducted under the proportional hazards assumption. In Section 4 we compute the mean and variance of the PLE, and corresponding upper and lower bounds of the PEXE. We compare our exact result for the bias of the PLE with Efron's (1967) upper bound. It is seen that Efron's bound, in the proportional hazards model, is quite loose. We also compare the exact variance of the PLE with an approximate variance given by Kaplan and Meier (1958). The latter approximation is found to be quite good.

It is also seen in Section 4 that our upper bound for the mean of the PEXE is reasonably close to the true value $\overline{F}(t)$ but the lower bound is not good. Furthermore, the upper bound for the variance of the PEXE is compared to the approximation given by Kaplan and Meier (1958) for the variance of the PLE. (Recall that the PEXE and the PLE have the same asymptotic variance, and thus the Kaplan-Meier approximation can also be viewed as an approximation to the (small-sample) variance of the PEXE.) The upper bound is seen to be too conservative, and tighter bounds are needed.

2. PRELIMINARIES

In this section we define the PEXE, the PLE, and the proportional hazards models. Then we present a characterization of these models.

First we introduce some notation. Let $\overline{K}_n(t) = \Sigma_{i=1}^n I(Z_i > t)$, $t\epsilon[0, \infty)$, be the number of observations in the sample at time t, and let $\tau(n) = \Sigma_{i=1}^n \delta_i$ be the total number of failures. Further, let $W_{(1)} < W_{(2)} \cdots < W_{(\tau(n))}$ denote the consecutive observed failures with $W_{(0)} \equiv 0$, and for $\tau(n) \geq 1$, let $r_i = [\int_{W_{(i-1)}}^{W_{(i)}} n\overline{K}_n(u)du]^{-1}$ be the number of failures per unit time in the interval $(W_{(i-1)}, W_{(i)}]$, $i = 1, \ldots, \tau(n)$.

We are ready to present the definitions of the PEXE and the PLE.

Definition 2.1. For $\tau(n) \geq 1$ let $\Delta_i = (W_{(i)} - W_{(i-1)})r_i$, $i = 1, \ldots, \tau(n)$. Then the Piecewise Exponential Estimator, (PEXE), denoted by $\overline{S}_n(t)$, is equal to 1 on the set $\{\tau(n) = 0 \text{ or } t\epsilon(-\infty, 0]\}$, is equal to $\exp\{-\Sigma_{j=1}^{i-1}\Delta_i - (t - W_{(i-1)})r_{(i)}\}$, on the set $\{\tau(n) \geq 1,\ t\epsilon(W_{(i-1)}, W_{(i)}]\}$, $i = 1, \ldots, \tau(n)$, and is equal to $\exp\{-\Sigma_{i=1}^{\tau(n)}\Delta_i\}$, on the set $\{\tau(n) \geq 1,\ t\epsilon(W_{(\tau(n))}, \infty)\}$.

Definition 2.2. Let $\overline{K}_n(t-) = n^{-1}\Sigma_{i=1}^n I(Z_i \geq t)$, $t\epsilon[0,\infty)$. Then the Product Limit Estimator, (PLE), denoted by $\overline{F}_n(t)$, is equal to 1 on the set $\{\tau(n) = 0, \text{ or } t\epsilon(-\infty,0)\}$, is equal to $\prod_{j=1}^{i-1}[\overline{K}_n(W_{(j)})][\overline{K}_n(W_{(j)}-)]^{-1}$, on the set $\{\tau(n) \geq 1,\ t\epsilon[W_{(i-1)}, W_{(i)})\}$, $i = 1, \ldots, \tau(n)$, and is equal to $\prod_{j=1}^{\tau(n)} [\overline{K}_n(W_{(j)})][\overline{K}_n(W_{(j)}-)]^{-1}$, on the set $\{\tau(n) \geq 1,\ t\epsilon[W_{(\tau(n))}, \infty)\}$.

We note that originally Kaplan and Meier (1958) left the PLE undetermined on the set $\{\tau(n) \geq 1,\ t\epsilon[\max_{1\leq i\leq n} Z_i, \infty)\}$. (See also Peterson (1977).)

Next we define a proportional hazards model. Let (U_1, U_2) be a pair of independent nonnegative r.v.'s with continuous d.f.'s G_1, G_2 respectively. Further, let $\overline{G}_q \equiv 1 - G_q$, $\alpha(G_q) = \sup\{t: G_q(t) < 1\}$, and let $R_q(t) = -\ln\overline{G}_q(t)$, $t\epsilon(-\infty, \alpha(G_q))$, be the hazard function corresponding to the d.f. G_q, $q = 1, 2$.

<u>Definition 2.3</u>. We say that the pair (U_1, U_2) follows a <u>proportional hazards model</u> if there is a positive real number β such that:

$$\bar{G}_2(t) = [\bar{G}_1(t)]^{\beta} \text{ for } t\in[0,\infty). \tag{2.1}$$

Note that from (2.1) $\alpha(G_1) = \alpha(G_2)$, and that the hazard functions corresponding to G_1, G_2 are proportional:

$$R_1(t) = \beta R_2(t),\ t\in[0,\ \alpha(G_1)).$$

Next we characterize the proportional hazards models.

<u>Theorem 2.4</u>. The pair (U_1, U_2), $0 < P\{U_1 \le U_2\} < 1$, follows a proportional hazards model if and only if

The r.v.'s $U = \min\{U_1, U_2\}$, and $\xi = I(U_1 \le U_2)$ are independent. (2.2)

<u>Proof</u>. Let G be a continuous d.f., let $\bar{G} \equiv 1 - G$, and assume that $G(0) = 0$. Then for $\theta\in[-1,\infty)$, and $z\in[0,\alpha(G))$:

$$\int_z^{\infty}[\bar{G}(u)]^{\theta}dG(u) = \begin{cases}[1+\theta]^{-1}[\bar{G}(z)]^{\theta+1}, & \theta > -1, \\ -\ln\bar{G}(z) & , \theta = -1.\end{cases} \tag{2.3}$$

First we prove that (2.1) implies (2.2). To verify (2.2) it suffices to show that for $z\in[0,\infty)$

$$P\{U > z,\ \xi = 1\} = P\{U > z\}P\{\xi = 1\}. \tag{2.4}$$

By (2.1) and (2.3)

$$\begin{aligned}P\{U > z,\ \xi = 1\} &= \int_z^{\infty}\bar{G}_2(u)dG_1(u) = \int_z^{\infty}[\bar{G}_1(u)]^{\beta}dG_1(u) \\ &= (1+\beta)^{-1}[\bar{G}_1(z)]^{\beta+1} = (1+\beta)^{-1}P\{U > z\}.\end{aligned} \tag{2.5}$$

In particular $P\{\xi = 1\} = P\{U > 0, \xi = 1\} = (1 + \beta)^{-1}$. Thus, (2.4) holds. Consequently (2.1) implies (2.2).

Now we show that (2.2) implies (2.1). Let $\gamma = P\{U_1 \leq U_2\}$. By (2.2):

$$\gamma\bar{G}_1(z)\bar{G}_2(z) = \int_z^\infty \bar{G}_2(u)dG_1(u), \ z\in[0,\infty). \tag{2.6}$$

Upon integration by parts

$$\gamma\int_z^\infty \bar{G}_1(u)dG_2(u) = (1 - \gamma)\int_z^\infty \bar{G}_2(u)dG_1(u), \ z\in[0,\infty). \tag{2.7}$$

Now (2.7) implies for $z\in[0, \min\{\alpha(G_1), \alpha(G_2)\})$, that:

$$\gamma\int_0^z [\bar{G}_2(u)]^{-1}dG_2(u) = (1 - \gamma)\int_0^z [\bar{G}_1(u)]^{-1}dG_1(u). \tag{2.8}$$

By (2.3) and (2.8), for $t\in[0,\min\{\alpha(G_1),\alpha(G_2)\})$,

$$\bar{G}_2(t) = [\bar{G}_1(t)]^{\gamma^{-1}(1-\gamma)}. \tag{2.9}$$

Consequently (2.2) implies (2.1). ||

Since the random pairs (X_i, Y_i), $i = 1, 2, \ldots$ are i.i.d. with independent components it follows that:

> The r.v.'s δ_i, $i = 1, 2, \ldots$ are i.i.d. Bernoulli with the parameter $\gamma = P\{X_1 \leq Y_1\}$. (2.10)

Further, under the assumption that (X_1, Y_1) follows a proportional hazards model we conclude from Theorem 2.4 that:

> The sequences of r.v.'s $\delta_1, \delta_2, \ldots$, and $Z_1, Z_2, \ldots$ are independent. (2.11)

Thus, the proportional hazards assumption yields a nice structure on the random pairs $(X_1, Y_1), (X_2, Y_2), \ldots$. This structure plays an important role in our computations, presented in Sections 3 and 4. Without this assumption finite sample computations are extremely difficult. (Thus our assumption is also pragmatic.)

3. MOMENT CALCULATIONS AND BOUNDS UNDER PROPORTIONAL HAZARDS.

In this section we present formulas for the mean and variance of the PLE. Then we give upper and lower bounds for the mean of the PEXE and an upper bound for the variance of the PEXE. These are derived under the assumption that (X_1, Y_1) follows a proportional hazards model. Throughout we define a sum and a product over an empty set of indices as 0 and 1 respectively.

First we present formulas for the mean and variance of the PLE. Let $K_n(t) = 1 - \overline{K}_n(t)$, $\overline{K}(t) = \overline{F}(t)\overline{H}(t)$, and $K(t) = 1 - \overline{K}(t)$, $t\epsilon[0,\infty)$. Note that by the continuity of F the formula for the PLE, given in Definition 2.2, reduces to:

$$\overline{F}_n(t) = \prod_{i=1}^{K_n(t)} [(n - i)(n - i + 1)^{-1}]^{\delta_i}, \quad t\epsilon(-\infty,\infty). \tag{3.1}$$

Let $t\epsilon[0,\infty)$. Then $K_n(t)$ is a binomial r.v. with parameters n and $K(t)$, $\delta_1, \ldots \delta_n$ are i.i.d. Bernoulli r.v.'s with parameter γ, and by (2.11) $K_n(t)$ is independent of $\delta_1, \ldots, \delta_n$. Thus, by (3.1), for t, $\alpha\epsilon(0,\infty)$:

$$\begin{aligned} E\{\overline{F}_n(t)\}^\alpha &= E \prod_{i=1}^{K_n(t)} [\gamma(n - i)^\alpha(n - i + 1)^{-\alpha} + 1 - \gamma] \\ &= \Sigma_{q=0}^n \binom{n}{q}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} \{\gamma(n - i)^\alpha(n - i + 1)^{-\alpha} + 1 - \gamma\}. \end{aligned}$$

Consequently for $t\epsilon[0,\infty)$:

$$E\overline{F}_n(t) = \Sigma_{q=0}^n \binom{n}{q}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} (1 - \gamma(n - i + 1)^{-1}), \tag{3.2}$$

and

$$\begin{aligned} Var\{\overline{F}_n(t)\} &= \Sigma_{q=0}^n \binom{n}{q}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} \{1 - \gamma(2n - 2i + 1)(n - i + 1)^{-2}\} \\ &\quad - \{\Sigma_{q=0}^n \binom{n}{q}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} (1 - \gamma(n - i + 1)^{-1})\}^2. \end{aligned} \tag{3.3}$$

Equations (3.2) and (3.3) are conveniently used for numerical comparisons in Section 4.

Now we obtain upper and lower bounds for the mean of the PEXE and an upper bound for the variance of the PEXE. Note that

$$e^{-x^{-1}} \leq x(1+x)^{-1} \leq e^{-(x+1)^{-1}}, \; x\epsilon(0,\infty), \tag{3.4}$$

and that on the set $\{\tau(n) \geq 1\}$

$$[n\overline{K}_n(W_{(i-1)})]^{-1} \leq \Delta_i \leq [n\overline{K}_n(W_{(i)})]^{-1}, \; i = 1, \ldots, \tau(n). \tag{3.5}$$

For $t\epsilon(0,\infty)$, let $a(n, t) = \max\{q: \; q = 0, \ldots, \tau(n), W_{(q)} \leq t\}$ be the index of the largest $W_{(q)}$ preceding t, and on the set $\{a(n, t) < \tau(n)\}$ let $b(n, t) = \min\{q: \; q = K_n(t) + 1, \ldots, n, \delta_{(q)} = 1\}$ be the index of the smallest observation that failed following t. By the formula for the PEXE, given in Definition (2.1), and by (3.4), (3.5):

$$\begin{aligned}
\overline{S}_n(t) &\leq \exp\{-\Sigma_{i=1}^{a(n,t)}\Delta_i\} \\
&\leq \exp\{-\Sigma_{i=1}^{a(n,t)}[n\overline{K}_n(W_{(i-1)})]^{-1}\} \\
&\leq \prod_{i=1}^{a(n,t)} \{[\overline{K}_n(W_{(i-1)})][\overline{K}_n(W_{(i-1)}) + n^{-1}]^{-}\} \\
&= \prod_{i=1}^{K_n(t)} \{(n-i+1)(n-i+2)^{-1}\}^{\delta_i} \text{ for } t\epsilon(0,\infty),
\end{aligned}$$

and

$$\overline{S}_n(t) \geq [I(a(n, t) = \tau(n)) + I(a(n, t) < \tau(n))\exp\{-\Delta_{b(n,t)}\}]\exp\{-\Sigma_{i=1}^{a(n,t)}\Delta_i\}$$

$$\geq [I(a(n, t) = \tau(n)) + I(a(n, t) < \tau(n))\exp\{-\Delta_{b(n,t)}\}]\exp\{-\Sigma_{i=1}^{a(n,t)}[n\overline{K}_n(W_{(i)})]^{-1}\}$$

$$\geq [I(a(n, t) = \tau(n)) + \tfrac{1}{2}I(a(n, t) < \tau(n), b(n, t) < n - 1]$$

$$I(K_n(t) < n) \prod_{i=1}^{K_n(t)} \{(n - i - 1)(n - i)^{-1}\}^{\delta_i}$$

$$\geq \tfrac{1}{2}I(K_n(t) < n) \prod_{i=1}^{K_n(t)} \{(n - i - 1)(n - i)^{-1}\}^{\delta_i} \text{ for } t\epsilon[0, \infty).$$

Thus, for t, $\alpha\epsilon(0,\infty)$

$$\begin{aligned} E\{\overline{S}_n(t)\}^\alpha &\leq E \prod_{i=1}^{K_n(t)} [\gamma(n - i + 1)^\alpha(n - i + 2)^{-\alpha} + 1 - \gamma] \\ &= \Sigma_{q=0}^{n}\binom{n}{q}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} [\gamma(n - i + 1)^\alpha(n - i + 2)^{-\alpha} + 1 - \gamma], \end{aligned} \tag{3.6}$$

and

$$\begin{aligned} E\{\overline{S}_n(t)\}^\alpha &\geq 2^{-\alpha}E \prod_{i=1}^{K_n(t)} [\gamma(n - i - 1)^\alpha(n - i)^{-\alpha} + 1 - \gamma] \\ &= 2^{-\alpha}\Sigma_{q=0}^{n-1}\binom{n}{q}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} [\gamma(n - i - 1)^\alpha(n - i)^{-\alpha} + 1 - \gamma]. \end{aligned}$$

Consequently

$$E\overline{S}_n(t) \leq \Sigma_{q=0}^{n}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} (1 - \gamma(n - i + 2)^{-1}) \tag{3.7}$$

$$E\overline{S}_n(t) \geq 2^{-1}\Sigma_{q=0}^{n-1}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} (1 - \gamma(n - i)^{-1}) \tag{3.8}$$

and

$$\begin{aligned} \mathrm{Var}\{\overline{S}_n(t)\} &\leq \Sigma_{q=0}^{n}\binom{n}{q}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} (1 - \gamma(2n - 2i + 3) \\ &(n - i + 2)^{-2}) - \{2^{-1}\Sigma_{q=0}^{n-1}\binom{n}{q}[K(t)]^q[\overline{K}(t)]^{n-q} \prod_{i=1}^{q} (1 - \gamma(n - i)^{-1})\}^2. \end{aligned} \tag{3.9}$$

4. COMPARISON OF EXACT VALUES WITH BOUNDS AND APPROXIMATE VALUES

In this section we take the lifelength d.f. F to be exponential with scale parameter 1 and the right-censorship d.f. H to be exponential with scale parameter $\beta(\beta > 0)$.

Table 1 displays results obtained by using (3.2) and (3.3) to obtain numerical values for the mean and variance of the PLE. Then we compare the bias of the PLE to the general bound for the bias given by Efron (1967), namely:

$$0 \leq E\overline{F}_n(t) - \overline{F}(t) \leq F(t)\exp\{-n\overline{K}(t)\}, \; t\epsilon[0,\infty). \tag{4.1}$$

We also compare the exact variance of the PLE to the approximate variance given by Kaplan and Meier (1958):

$$\mathrm{Var}\{\overline{F}_n(t)\} \approx n^{-1}[\overline{F}(t)]^2\int_0^t[\overline{K}(u)\overline{F}(u)]^{-1}dF(u), \; t\epsilon[0,\infty). \tag{4.2}$$

We note that $[\overline{F}(t)]^2\int_0^t[\overline{K}(u)\overline{F}(u)]^{-1}dF(u)$ is the variance of $G(t)$, the Gaussian process obtained as the weak limit as $n \to \infty$ of $n^{1/2}\{\overline{F}_n(t) - \overline{F}(t)\}$. $G(t)$ is also the weak limit of $n^{1/2}\{\overline{S}_n(t) - \overline{F}(t)\}$. [See KLP (1980b).] Thus the right-hand side of (4.2) can also be considered as an approximation to $\mathrm{Var}\{\overline{S}_n(t)\}$.

We also use (3.7), (3.8), and (3.9) to obtain lower and upper bounds for the mean of the PEXE, and upper bounds for the variance of the PEXE. See Table 2.

All our calculations are performed under a particular proportional hazards model, when X and Y are exponential r.v.'s. However these calculations apply to general proportional hazards models, provided the appropriate adjustments are made. More specifically assume

$$\bar{H} = [\bar{F}]^{\beta} \text{ for some } \beta > 0, \tag{4.3}$$

and let $R^{-1}(t) = \inf\{z: z \in [0, \alpha(F)), -\ln \bar{F}(z) > t\}$, $t \in [0, \infty)$. Since F is continuous $R(R^{-1}(t)) = t$ for $t \in [0, \infty)$. Thus:

The r.v.'s $R(X)$ and $R(Y)$ are exponentially distributed with scale parameters 1 and β respectively.

Let $Z_{i,R} = \min\{R(X_i), R(Y_i)\}$ and $\delta_{i,R} = I(R(X_i) \leq R(Y_i))$, $i = 1, \ldots, n$. Then:

The random vectors $\langle (Z_{(1),R}, \delta_{(1),R}), \ldots, (Z_{(n),R}, \delta_{(n),R}) \rangle$ and $\langle (R(Z_{(1)}), \delta_{(1)}), \ldots, (R(Z_{(n)}), \delta_{(n)}) \rangle$ are stochastically equal.

Now let $\bar{F}_{n,R}$ and $\bar{S}_{n,R}$ be the PLE and PEXE computed from $(Z_{(1),R}, \delta_{(1),R}), \ldots,$ $(Z_{(n),R}, \delta_{(n),R})$, respectively. Then for $n = 1, 2, \ldots$:

The two PLE's processes $\{\bar{F}_{n,R}(R(t)), t \in [0, \infty)\}$ and $\{\bar{F}_n(t), t \in [0, \infty)\}$ are stochastically equal.

In this section we compare the exact bias and exact variance of the PLE, for proportional hazards models, with Efron's (1967) upper bound for the bias and with Kaplan and Meier's (1958) approximate variance. Thus, we may without loss of generality replace the model, given by (4.3), with the particular exponential model described in the first paragraph of this section.

We note that the PEXE processes $\{\bar{S}_{n,R}(R(t)), t \in [0, \infty)\}$ and $\{\bar{S}_n(t), t \in [0, \infty)\}$ are not stochastically equal. However, since the upper and lower bounds of $\bar{S}_n$, given in Section 3, depend only on $K_n(t)$ and $\tau(n)$, our comparisons for the PEXE conducted for an exponential model apply to a general proportional hazards model provided t is replaced by $R(t)$.

Next we summarize our results for the PLE in the following table.

1. Exact and Approximate Values of the Mean and Variance of the PLE

$\beta = 0.5,\ t = 0.5,\ \bar{F}(t) = .6065$

sample size n	10	15	20	25	30
mean of the PLE	.6065	.6065	.6065	.6065	.6065
bias of the PLE	.0000	.0000	.0000	.0000	.0000
Efron's bound for the bias	.0035	.0003	.0000	.0000	.0000
variance of the PLE	.0284	.0185	.0138	.0110	.0092
approximate variance	.0274	.0183	.0137	.0110	.0091

$\beta = 1,\ t = 0.5,\ \bar{F}(t) = .6065$

sample size n	10	15	20	25	30
mean of the PLE	.6066	.6065	.6065	.6065	.6065
bias of the PLE	.0001	.0000	.0000	.0000	.0000
Efron's bound for the bias	.0099	.0016	.0003	.0000	.0000
variance of the PLE	.0333	.0218	.0162	.0129	.0107
approximate variance	.0316	.0211	.0158	.0126	.0105

$\beta = 2,\ t = 0.5,\ \bar{F}(t) = .6065$

sample size n	10	15	20	25	30
mean of the PLE	.6087	.6069	.6066	.6065	.6065
bias of the PLE	.0022	.0004	.0001	.0000	.0000
Efron's bound for the bias	.0423	.0138	.0045	.0015	.0005
variance of the PLE	.0478	.0313	.0229	.0181	.0149
approximate variance	.0427	.0285	.0213	.0171	.0142

$\beta = 0.5,\ t = 1,\quad \bar{F}(t) = .3679$

sample size n	10	15	20	25	30
mean of the PLE	.3688	.3680	.3679	.3679	.3679
bias of the PLE	.0009	.0002	.0000	.0000	.0000
Efron's bound for the bias	.0395	.0129	.0042	.0014	.0005
variance of the PLE	.0334	.0219	.0129	.0107	.0107
approximate variance	.0314	.0209	.0157	.0126	.0105

$\beta = 1,\ t = 1,\ \bar{F}(t) = .3679$

sample size n	10	15	20	25	30
mean of the PLE	.3752	.3701	.3686	.3682	.3680
bias of the PLE	.0073	.0022	.0008	.0003	.0001
Efron's bound for the bias	.1633	.0830	.0422	.0214	.0109
variance of the PLE	.0474	,0318	.0236	.0173	.0144
approximate variance	.0423	.0288	.0216	.0186	.0154

$\beta = 2,\ t = 1,\ \bar{F}(t) = .3679$

sample size n	10	15	20	25	30
mean of the PLE	.4164	.3946	.3836	.3799	.3744
bias of the PLE	.0485	.0268	.0160	.0100	.0065
Efron's bound for the bias	.3842	.2995	.2335	.1821	.1419
variance of the PLE	.0779	.0580	.0462	.0382	.0324
approximate variance	.0861	.0574	.0430	.0344	.0287

$\beta = 0.5,\ t = 2,\ \bar{F}(t) = .1353$

sample size n	10	15	20	25	30
mean of the PLE	.1546	.1449	.1406	.1384	.1373
bias of the PLE	.0193	.0095	.0053	.0031	.0019
Efron's bound for the bias	.5256	.4097	.3194	.2491	.1942
variance of the PLE	.0254	.0169	.0127	.0101	.0084
approximate variance	.0233	.0155	.0117	.0093	.0078

$\beta = 1,\ t = 2,\ \bar{F}(t) = .1353$

sample size n	10	15	20	25	30
mean of the PLE	.2076	.1825	.1688	.1602	.1544
bias of the PLE	.0723	.0472	.0335	.0249	.0192
Efron's bound for the bias	.7200	.6569	.5995	.5470	.4991
variance of the PLE	.0470	.0327	.0252	.0205	.0174
approximate variance	.0491	.0327	.0245	.0196	.0164

$\beta = 2,\ t = 2,\ \bar{F}(t) = .1353$

sample size n	10	15	20	25	30
mean of the PLE	.3432	.3027	.2772	.2591	.2455
bias of the PLE	.2079	.1674	.1479	.1238	.1101
Efron's bound for the bias	.8435	.8331	.8228	.8127	.8027
variance of the PLE	.0861	.0672	.0560	.0485	.0431
approximate variance	.2457	.1638	.1228	.0983	.0819

We see that Efron's (1967) bound for the bias is much larger than the exact bias, computed for proportional hazards models. However the approximate variance given by Kaplan and Meier (1958) is close to the exact variance, even for the smallest sample size considered ($n = 10$).

The bounds for the PEXE in Table 2 are based on formulas (3.7), (3.8) and (3.9). The upper bound for the mean of the PEXE is reasonably close to the true value of $\bar{F}(t)$ but the lower bound for the mean is not close. Sharper bounds need to be developed.

For completeness, we include the upper bound for the variance of the PEXE but until an exact expression is obtained for the (small-sample) variance of the PEXE, the potential value and use of this upper bound is diminished. However, we can compare the upper bound to the approximate variance given by the right-hand side of (4.2). Such a comparison indicates that the upper bound is not that good.

2. PEXE Bounds: Upper and Lower Bounds for the Mean and Upper Bound for the Variance

$\beta = 0.5$, $t = 0.5$, $\bar{F}(t) = .6065$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.6476	.6348	.6280	.6239	.6211
lower bound for the mean of the PEXE	.2768	.2866	.2911	.2937	.2954
upper bound for the variance of the PEXE	.3649	.3366	.3220	.3131	.3070
approximate variance	.0274	.0183	.0137	.0110	.0091

$\beta = 1,\ t = 0.5,\ \bar{F}(t) = .6065$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.6539	.6391	.6313	.6266	.6233
lower bound for the mean of the PEXE	.2706	.2831	.2888	.2919	.2940
upper bound for the variance of the PEXE	.3792	.3463	.3292	.3188	.3119
approximate variance	.0316	.0211	.0158	.0126	.0105

$\beta = 2,\ t = 0.5,\ \bar{F}(t) = .6065$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.6700	.6505	.6400	.6336	.6292
lower bound for the mean of the PEXE	.2510	.2712	.2810	.2864	.2898
upper bound for the variance of the PEXE	.4162	.3723	.3488	.3334	.3247
approximate variance	.0427	.0285	.0213	.0171	.0142

$\beta = 0.5,\ t = 1,\ \bar{F}(t) = .3679$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.4453	.4212	.4085	.4007	.3954
lower bound for the mean of the PEXE	.1352	.1510	.1597	.1650	.1685
upper bound for the variance of the PEXE	.2044	.1723	.1552	.1447	.1376
approximate variance	.0314	.0209	.0157	.0126	.0105

$\beta = 1,\ t = 1,\ \bar{F}(t) = .3679$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.4722	.4407	.4237	.4130	.4058
lower bound for the mean of the PEXE	.1192	.1365	.1477	.1553	.1606
upper bound for the variance of the PEXE	.2386	.1980	.1755	.1613	.1516
approximate variance	.0432	.0288	.0216	.0186	.0154

$\beta = 2,\ t = 1,\ \bar{F}(t) = .3679$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.5396	.4966	.4707	.4532	.4407
lower bound for the mean of the PEXE	.0783	.0981	.1097	.1205	.1292
upper bound for the variance of the PEXE	.3247	.2694	.2364	.2142	.1282
approximate variance	.0861	.0574	.0430	.0344	.0287

$\beta = 0.5,\ t = 2,\ \bar{F}(t) = .1353$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.2782	.2375	.2148	.2003	.1902
lower bound for the mean of the PEXE	.0251	.0298	.0339	.0374	.0404
upper bound for the variance of the PEXE	.0938	.0676	.0544	.0465	.0412
approximate variance	.0223	.0155	.0117	.0093	.0078

$\beta = 1,\ t = 2,\ \bar{F}(t) = .1353$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.3566	.3049	.2737	.2525	.2370
lower bound for the mean of the PEXE	.0170	.0205	.0234	.0260	.0283
upper bound for the variance of the PEXE	.1538	.1123	.0901	.0762	.0667
approximate variance	.0491	.0327	.0245	.0196	.0164

$\beta = 2,\ t = 2,\ \bar{F}(t) = .1353$

sample size n	10	15	20	25	30
upper bound for the mean of the PEXE	.4955	.4398	.4033	.3769	.3565
lower bound for the mean of the PEXE	.0043	.0056	.0067	.0078	.0087
upper bound for the variance of the PEXE	.2869	.2272	.1915	.1674	.1499
approximate variance	.2457	.1638	.1228	.0983	.0819

REFERENCES

[1] Breslow, N. and Crowley T. (1974). A large sample study of the life table and product limit estimates under random censoring Ann. Statist. 2 437-453.

[2] Efron, B. (1967). The two sample problem with censored data. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. 831-852.

[3] Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. J. Amer. Statist. Assoc. 53 457-481.

[4] Kitchin, J., Langberg, N. A. and Proschan F. (1980a). A new method for estimating life distributions from incomplete data. Florida State University Report M548.

[5] ______________________________ (1980b). The weak convergence of the piecewise exponential estimater. Florida State University Report M551.

[6] Langberg, N. A., Proschan, F., and Quinzi, A. T. (1980) Estimating dependent life lengths, with applications to the theory of competing risks. Ann. Statist. 8.

[7] Peterson, A. V. (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. J. Amer. Statist. Assoc. 72 854-858.